A Model for Designing Adaptive Laboratory Evolution Experiments

LaCroix, Ryan A.; Palsson, Bernhard O.; Feist, Adam M.; Kivisaar, Maia

Published in:
Applied and Environmental Microbiology

Link to article, DOI:
10.1128/AEM.03115-16

Publication date:
2017

Document Version
Peer reviewed version

Appl. Environ. Microbiol. doi:10.1128/AEM.03115-16

A Model for Designing Adaptive Laboratory Evolution Experiments

Ryan A. LaCroix[1], Bernhard O. Palsson[1,2,3], Adam M. Feist[1,2]

[1] Department of Bioengineering, University of California, San Diego, California, United States

[2] Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Lyngby, Denmark

[3] Department of Pediatrics, University of California, San Diego, California, United States

Running Head: Designing Adaptive Laboratory Evolution Experiments

Address correspondence to Adam M. Feist, afeist@ucsd.edu

Abstract

The occurrence of mutations is a cornerstone of the evolutionary theory of adaptation, capitalizing on the rare chance that a mutation confers a fitness benefit. Natural selection is increasingly being leveraged in laboratory settings for industrial and basic science applications. Despite an increasing deployment, there are no standardized procedures available for designing and performing adaptive laboratory evolution (ALE) experiments. Thus, there is a need to optimize the experimental design, specifically for determining when to consider an experiment complete and for balancing outcomes with available resources (i.e., lab supplies, personnel, and time). To design and better understand ALE experiments, a simulator, ALEsim, was developed, validated, and applied to optimize ALE experimentation. The effects of various passage sizes were experimentally determined and subsequently evaluated with ALEsim to explain differences in experimental outcomes. Further, a beneficial mutation rate of $10^{-6.9}$-$10^{-8.4}$ mutations per cell division was derived. A retrospective analysis of ALE experiments revealed that passage sizes typically employed in serial passage batch culture ALE experiments led to inefficient production and fixation of beneficial mutations. ALEsim and the results herein will aid in the design of ALE experiments to fit the exact needs of the project while taking into account the tradeoff in resources required, and lower the barrier of entry to this experimental technique.

Importance

Adaptive laboratory evolution (ALE) is a widely used scientific technique to increase scientific understanding, as well as create industrially relevant organisms. The manner in which ALE experiments are conducted is highly manual and uniform with little optimization for efficiency. Such inefficiencies result is a suboptimal experiments that can take multiple months to complete. With the availability of automation and computer simulations, we can now perform these experiments in a more optimized fashion and design experiments to generate greater fitness in a more accelerated time frame, thereby pushing the limits of what adaptive laboratory evolution can achieve.

Highlights

- A tunable simulator, ALEsim, was constructed to simulate observed fitness increases in ALE experiments
- A control ALE experiment was performed to determine an observed beneficial mutation rate and quantify the effect of passage size in an ALE experiment – the beneficial mutation rate (BMR) is consistent with previous estimates
- A retrospective analysis of ALE experiments revealed limitations in experimental designs.
- ALEsim can be leveraged to optimize resources and time needed to conduct an ALE experiment by determining tradeoffs between a likely fitness increase and an increased run time

Introduction

Adaptive laboratory evolution (ALE) has been performed *in vitro* for decades and the field is expanding. ALE involves subjecting a population of organisms to a given environment, in the lab, and allowing natural selection to increase the overall fitness of the population. In laboratory settings, this is typically performed with organisms possessing short generation times. The basic principles governing ALE experiments are easily understood across a breadth of disciplines, which has led to its adoption in many laboratories (1, 2). The recent growth in the use of ALE can be attributed to the ease of access and decreasing costs of genome sequencing (3-5). Falling sequencing costs have led to the increased investigation of genomic, transcriptomic, and additional omics data types over the course of evolution (5). While the analysis of ALE experiments has grown, the manner in which the ALE experiments themselves are performed has remained relatively *ad hoc*. The most commonly employed techniques are chemostat adaptation and serially passaged batch culture adaptation, with batch culture adaptation being more popular as it is easily expanded and does not require setting up complex machinery (3, 6).

A primary attribute of any ALE experiment is the selection pressure imposed on the culture. The selection pressure (i.e., exponential growth, biomass yield, stationary phase, or lag phase) is responsible for the outcome of the evolution study (4, 7-10). For example, in a 24hr serially passaged batch culture ALE experiment with fast growing bacteria, the culture is subjected to alternating environments of feast and famine. At the beginning of each batch there are excess nutrients but inevitably, within 24hrs, the nutrients are consumed and stationary phase is reached. Because of this alternating environment, the selection pressure is complex and fitness is achieved through various methods (e.g., stationary phase fitness, lag phase duration, and growth rate all contribute) (9). This complexity often confounds the analysis depending on the application. To

alleviate complexity, the cells can be kept in one phase (e.g. exponential phase) to mitigate most of the alternating selection and focus selection specifically on fitness gains through growth rate. In such cases, fitness will be treated as interchangeable with growth rate. The desired outcome of the experiment would dictate the ideal selection pressure to be imposed and thereby the experimental design, but the difference between the two designs is non trivial.

There are several parameters that affect the outcome of a serially passaged batch culture ALE experiment. A primary parameter involved is the passage size (11-13). Specifically, passage size determines how much of the population is allowed to propagate to each subsequent batch culture. If a beneficial mutation occurs, but is lost when the bottleneck is imposed, the rate of evolution can be slowed or even halted. Since smaller passage sizes can hinder the rate of evolution, it is often easier to perform a serially passaged batch culture ALE under alternating environments of feast and famine where a change in passage size only effects the duration of growth and stationary phases. However, if the application requires exponential phase passaging, a change in passage size also changes the time when the culture must be passaged. Because of this, the passage size is often dictated by an individual's schedule. Typically, the time in between passaging can be no shorter than ~12hrs. Consequentially, as the culture adapts and begins to grow faster, the passage size must be decreased. As an example, a previous study adapting *E. coli* to glycerol in 250mL batches started with a passage size of approximately 100µL and by experiment's end was less than 0.1µL (14). A more in-depth retrospective analysis revealed similar trends where passage amounts were significantly decreased (14-18). In these studies, the reduction in population size, or bottleneck, (i.e., passage size) became so significant that the calculated number of cells being passed was on the order of 10 or even occasionally 1. The chance of capturing a beneficial mutation, when only passing tens of cells from a culture of

Applied and Environmental Microbiology

98 millions, is practically null over a reasonable timeframe. At this point, continuing the experiment
99 is futile. The question then becomes at what point is the passage size too low?

100 Passage size can have a large impact on the trajectory of an ALE experiment. This can be seen in
101 the comparison of two studies that evolved wild-type *E. coli* K-12 MG1655 on M9 glucose
102 minimal media (7, 18). One study (7) used a consistent passage size of 800µL from 25mL
103 batches on an automated platform. The second study (18) was done "by hand" and had widely
104 varying passage sizes that were considerably smaller than the automated study. The outcomes of
105 the ALE experiments were quite distinct. The final growth rates achieved were 1.00±0.24 hr^{-1}
106 and 0.79±.01 hr^{-1} in the consistent and variable passage size studies, respectively. The apparent
107 lack of fitness achieved in variable passage study was not due to a lack of available beneficial
108 mutations (as the same strains and culturing conditions were used), but rather insufficient
109 experimental design to find and fix them in a reasonable amount of time. Understanding why
110 these two outcomes differ is imperative to the efficient design of ALE experiments.

111 Theoretical studies have looked at the effect of passage size on serially passaged batch culture
112 adaptation and resulted in varying predictions of an ideal passage size depending on the model
113 used (19, 20). The ideal passage sizes calculated are ideal from a mathematical standpoint. This
114 essentially gives the best chance for various mutations of different selective advantages to fix in
115 a population. The ideal passage sizes calculated in these studies are relatively large (13.5% and
116 20%)(19, 20). As mentioned previously, a larger passage necessitates an increase in resources.
117 More specifically, the resources required increase exponentially with passage size, yet the gains
118 slowly diminish. This work thus focuses on examining the diminishing returns in the context of
119 the desired result and the resources available. We set out to examine the impact of the key ALE
120 parameter: passage size. To address this, we created an *in silico* evolutionary model that

simulates the dynamics of capturing and fixing beneficial mutations in the context of an exponentially-passed batch culture ALE experiment. After building the model, we parameterize it using a combination of 30 independent ALE experiments of *E. coli* on glycerol minimal media across five different passage sizes (10%, 1%, 0.1%, 0.01%, and 0.001%). Using the parameterized model, we investigated the biological consequences of changing passage sizes and how close to optimal a given experiment is. With this knowledge, an experiment can be designed to fit the desired outcome, giving consideration to the resources required to achieve it, and the feasibility of performing such an experiment.

Materials and Methods

Adaptive Laboratory Evolution

Adaptive laboratory evolutions were started from wild-type *E. coli* strain MG1655 (ATCC47076) glycerol frozen stock and grown up overnight in 15mL magnetically stirred 0.2% glycerol M9 minimal media supplemented with trace elements. The magnet was stirred at 1150rpm, sufficient for completely aerobic growth. 30 experiments were started from 150µL aliquots from the overnight pre-culture. The experiments were subsequently grown in identical vessels and media as the pre-culture. Culture optical densities at 600nm (OD) were monitored over the course of each batch culture. When the culture reached an OD of 0.300 (±10%) as measured by a plate-reader with 100µL sample volume in a 96 well flat bottom microplate, an aliquot was taken and passed to a new batch culture filled with sterile media. An OD of 0.300 was chosen to preclude reaching stationary phase in any of the cultures and ensures OD measurements have not begun to saturate. Growth rates of each culture were determined using OD measurements taken over the lifetime of each batch culture.

Media

All cultures were grown in 0.2% glycerol M9 minimal media. The media consisted of 0.2% glycerol by volume, 0.1mM $CaCl_2$, 2.0mM $MgSO_4$, Trace element solution and M9 salts. 4000X Trace element solution consisted of 27g/L $FeCl_3*6H_2O$, 2g/L $ZnCl_2*4H_2O$, 2g/L $CoCl_2*6H_2O$, 2g/L $NaMoO_4*2H_2O$, 1g/L $CaCl_2*H_2O$, 1.3g/L $CuCl_2*6H_2O$, 0.5g/L H_3BO_3, and Concentrated HCl dissolved in ddH_2O and sterile filtered. 10x M9 Salts solution consisted of 68g/L Na_2HPO_4 anhydrous, 30g/L KH_2PO_4, 5g/L NaCl, and 10g/L NH_4Cl dissolved ddH_2O and autoclaved. Final concentrations in the media were 1x.

152 **DNA Sequencing**

153 Genomic DNA was isolated using Macherey-Nagel NucleoSpin® Tissue kit. The quality of
154 DNA was assessed with UV absorbance ratios using a Nano drop. DNA was quantified using
155 Qubit dsDNA High Sensitivity assay. Paired-end resequencing libraries were generated using
156 Illumina's Nextera XT kit with 700 pg of input DNA total. Sequences were obtained using an
157 Illumina Miseq with a MiSeq 600 cycle reagent kit v3. The breseq pipeline version 0.23 with
158 bowtie2 was used to map sequencing reads and identify mutations relative to the *E. Coli* K12
159 MG1655 genome (NCBI accession NC_000913.2) (21). All samples had an average mapped
160 coverage of at least 25x.

161 **Computer Modeling**

162 Modeling of simulations was computed using MATLAB 2015b on a Windows 7 professional
163 platform. Detailed descriptions are found as comments in the supplemental m-files. The
164 beneficial mutation rate was computed by a maximum likelihood estimation. It was calculated
165 for making a transition from State 1 to State 2 and State 2 to State 3 for passage sizes of 0.01%
166 and 0.001%. These passage size were chosen as they were the only ones that showed a
167 distribution of states achieved. The transition from State 1 to State 2 was capped at 20 days to
168 give a maximally distributed data set. The transition from State 2 to State 3 was started by
169 assuming that State 2 was already achieved. Thus, the length of time simulated was started based
170 of when State 2 was achieved. This was variable for different experiments.

171 A value of 1.55×10^{12} cells$\cdot L^{-1} \cdot OD_{600nm}{}^{-1}$ was used to estimate the number of cells in a culture
172 for a given OD_{600nm} with a 1 cm path length cuvette for the purposes of ALEsim. A standard
173 curve relating the ODs measured in the plate reader with a 100µL sample volume in a 96 well

174 flat bottom microplate to the OD measured with a 1 cm cuvette to obtain a ratio of 3.15 for
175 equivalent measurements between the two. The biomass (grams of dry weight) per OD_{600nm} per
176 volume was calculated by filtering known volumes of cultures at specific ODs though 0.22µm
177 filters. The filters were weighed before and after filtering and drying to obtain the total dry
178 weight of the culture. The differences in these values was used to calculate ratio of $0.45 \cdot gDW\ L^{-1} \cdot OD_{600nm}^{-1}$.
179 The dry mass per cell has previously been reported as $2.9 \times 10^{-13} gDW \cdot cell$ (22). The
180 quotient of these two values gives our final conversion factor of 1.55×10^{12} $cells \cdot L^{-1} \cdot OD^{-1}$ to
181 estimate the cell counts of cultures at various ODs and volumes. . For *E. coli,* the dry mass per
182 cell can vary over a range of growth rates (23). Using such a variable OD to cell count factor as a
183 function of growth rate is possible with ALEsim, but incurs a marked increase in simulation
184 time. Thus, identical simulations were performed using only the highest and lowest dry mass per
185 cell values expected for the growth rates observed (i.e., the extremes). Only a 10% difference in
186 the distribution of simulated endpoint growth rates were observed between the two extremes (see
187 Supplementary Figure S1). Therefore, use of a constant average value for dry mass per cell over
188 the range of growth rates expected was determined to be sufficient considering the benefit in
189 computation time.

190 Although possible with ALEsim, deleterious and neutral mutations were not considered during
191 this study. A deleterious mutation rate of 1 in 5,000 was previously computed (24). In the
192 application demonstrated here, the population sizes were sufficiently large (10^5 – 10^9 cells) such
193 that the effects of deleterious and neutral mutations would be negligible. With smaller population
194 sizes (e.g., several orders of magnitude smaller than the population sizes modeled here), the
195 effects of these mutations become more pronounced and should not be ignored.

196

Results

Modeling the ALE process

ALEsim is a model built on the basic principles of exponential growth in order to understand the dynamics of ALE. The scope of ALEsim is to predict the observed growth rate in each batch culture of an ALE experiment while allowing individual cells to change their growth rate when dividing (i.e., a proxy for receiving a beneficial mutation). This preferentially finds only those beneficial mutations that fix. There is a likely chance that other beneficially mutations are unobserved due to clonal interference. The observed population growth rate is different from a clonal growth rate in that each batch culture of an ALE experiment is a population of multiple clones with varying growth rates. Figure 1 provides a workflow of the modeling process and the full details are in Supplementary File ALEsim.txt. Each *in silico* experiment begins with a clonal inoculation of a strain with a given growth rate. A population of mixed phenotypes can be used in this framework, but here the starting population will be assumed to be isogenic with the same phenotypic behavior. This organism is allowed to replicate according to an exponential growth function. During each cell division event, there is a probability that it will mutate and start a new lineage with a mutated growth rate. This new lineage is allowed to grow alongside the parent strain according to exponential growth, but with its mutated growth rate. The new lineage is itself allowed to continue mutating in the simulation.

Mutated growth rates in ALEsim must be constrained to remain biologically meaningful, i.e., growth rates that are of magnitudes that remain plausible. These rates are determined empirically by the user, as done here from the parameterization experiment (see section below). The growth rates can be constrained to allow various types of epistasis. For example, if two distinct growth rates are allowed, there is a possibility that a single cell line could mutate twice and receive both

of these mutations. ALEsim employs the flexibility to define the type of epistasis between these two mutations, if any epistasis at all is to occur. Similarly, an order to the mutations accumulated can be set, as certain mutations can be beneficial only in the presence of a pre-existing mutation (i.e., epistasis can be modeled). As the population of cells continues to replicate and mutate, their total cell count naturally increases. When the cell count reaches a given threshold, a simple random sample of cells is used to inoculate the next batch culture. The threshold corresponds to a target cell count at which to passage the cells to the next batch culture. The number of cells taken is determined by the passage size, which is a percentage of the total culture volume. After this sample is computed, a new batch culture is started with the chosen cells and corresponding growth rates. Figure 2 provides the key parameters of the model.

In using the basic principles of microbial growth and a brute force computational approach, many of the fundamental attributes of natural selection are intrinsically contained in the simulation. This includes clonal interference which is pervasive to asexual evolution. ALEsim can be used to model a system where two local maxima are possible but the greater maximum can only be found by first acquiring a mutation that is initially suboptimal compared to other possible single beneficial mutations (25). How to achieve this is shown in the model documentation (ALEsim.txt). The experimental parameters can be modulated to potentially find an experiment design that would find the desired optimum or both.

Given the stochastic nature of many steps in the model, the results are non-deterministic. Stochasticity is incorporated into the model in three ways: i) when a cell mutates its growth rate, ii) what growth rate a cell mutates to, and iii) what sample of cells are propagated to a subsequent batch culture. The simulation is then run multiple times to capture the dynamics of the stochasticity (26).

243 For a simulation to be biologically meaningful using the developed model, there are three types
244 of parameter sets that must be determined. The first set of parameters is experimental: batch
245 culture size, passage size, passage optical density (or cell count), and length of experiment.
246 These can be set based on the desired experimental setup.(23) The second set is the statistical
247 parameters: random number seed and the number of identical experiments to run. The random
248 number seed is set by the native random number generator. The number of parallel simulations to
249 run is determined by the statistical power needed. Depending on the magnitudes and
250 complexities of the parameters set, the number of simulations can vary drastically. For the results
251 shown here, 500 simulations were computed unless otherwise stated. It was found that after 500
252 simulations there was no appreciable difference in the means or spread of the distribution of
253 results calculated when combined with another set of 500. The third set of parameters is
254 biological: beneficial mutation rate (BMR) and allowed increases in growth rate. These
255 parameters are defined in the models and can be constrained by any method that can be
256 expressed programmatically, whether this it is randomly decided within a meaningful range or
257 set to distinct values. This set of parameters must be derived experimentally. Intuitively, these
258 parameters can be different for different strains, conditions, and can even change along the
259 course of a single experiment (27, 28). As long as the values determined are biologically
260 meaningful, generalizations about the ALE process can be concluded.

261 Alternative models of evolution and adaption have been developed to understand the dynamics
262 of evolution. These types of mathematical models capture various aspects of adaptation
263 including selection, drift, and clonal interference (29-31). Classically, this has been a target of
264 the field of population genetics (32-34). An expansion of the Fisher model was developed by
265 Wahl et. al. which conceptually relates to ALEsim in that it targets the question of passage sizes

(35). However, ALEsim deviates from the classical mathematical approach and employs the use of an *in silico* organism that can then replicate, mutate, and evolve. Simulations here are carried out in brute force where they are allowed to grow under the conditions laid out by the user. The advantage of such a method is that the experimental and biological parameters can be strictly controlled over the course of an experiment. The resulting simulation is able to more closely mimic the conditions of an actual laboratory evolution experiment in its entirety where parameters are not always constant throughout. This approach differs from the use of a digital organism in that it is an attempt to model specific biology instead of general evolutionary dynamics which allows for direct modeling of the ALE experiment as would be performed in a laboratory (36).

Parameterization of ALEsim by evolving *E. coli* on Glycerol Minimal Media

The two biological parameters, the beneficial mutation rate and allowed increase in growth rate, were determined using 30 independent cultures of *Escherichia coli* K-12 MG1655 evolved in 15mL of 0.2% glycerol M9 minimal media until a stable growth rate was observed in most experiments (38 days). One experiment only lasted 23 days after it was restarted due to contamination. The 30 experiments were separated into five groups of six passage sizes and each group was evolved under identical conditions except for the passage size. The passage sizes used were 10%, 1%, 0.1%, 0.01%, and 0.001% of the culture size (15mL). The growth rate of each experiment was monitored over the course of the experiment using optical density measurements as a proxy for cell count (Figure 3). Fitness related details can be found in the supplement (Supplementary Table 1 and Supplementary File fitness_data.xlsx).

Allowed increases in growth rate were determined by identifying jumps in growth rates from the fitness trajectories. A spline was fit to the growth rate of each experiment and significant

increases in growth rate were identified as discussed previously (7). The resulting jumps in growth rates showed that the plateaus in growth occurred at specific values (Figure 3, 4). These plateaus are identified as State 1, 2, 3A, and 3B. State 3 was split into two sub-states since there is a significant difference between those in state 3A and 3B (Wilcoxon rank sum $p<0.01$), however there exists no identifiable increase in growth rate or gap between states that would characterize this transition. This gap is most likely obscured since the difference between the growth rates is fairly small and noise in the measurements can bleed into any gap that might exist. Figure 4 groups the jumps in fitness observed by their transition between states. Contrary to the conclusion of other ALE experiments, the largest jump in fitness was not observed first but actually followed a smaller jump. This yields an allowed increase in growth rate that can be used to constrain ALEsim. In simulations run here, the growth rates allowed were set to the mean of the range of each state.

The beneficial mutation rate (BMR) can be calculated by fitting ALEsim to the distribution of the end states. Passage sizes of 10% - 0.1% did not show any appreciable variation between states, thus only the experiments with passage sizes of 0.01% and 0.001% were used for fitting. ALEsim was fit by performing simulations that only allowed for a single jump from one state to another. Multi-state jumps and two sequential jumps were not allowed. This simplification skews the BMR calculation to only include beneficial mutations that were fixed in the population. There is a potential that other beneficial mutations are possible, but were not observed due to either clonal interference or genetic drift (37). As observed in the fitness trajectories for passage sizes of 0.01% and 0.001%, not all experiments were able to make jumps to occupy all the states. For instance, with a passage size of 0.01%, only 4 of 6 experiments were able to make the transition from State 2 to State 3 by experiment's end. In simulation, the same distribution

among the various end states is observed. The distribution observed in simulation is highly dependent on the supply of beneficial mutations captured by the BMR parameter. Thus, the BMR can be fit to yield the same distribution across states as observed experimentally. The BMR was computed using transitions from both State 2 to State 3 and from State 1 to State 2. Since all experiments made the transition from State 1 to State 2, the distribution was used at the day 20 mark where a distribution existed. The 95% confidence interval for the BMR was calculated by fitting the BMR to the 95% confidence interval of the experimental distribution of states. The results yielded a BMR of $10^{-6.9}$-$10^{-8.4}$ mutations per cell division. The confidence interval was determined by a maximum likelihood estimate as implemented in the binofit function in MATLAB.

Retrospective Validation of ALEsim

ALEsim and the derived parameters (beneficial mutation rate and allowed increases in growth rate) were analyzed using two previously performed ALE experiments on glucose (7, 18) and a legacy experiment on glycerol (14). The outcomes of the two glucose experiments yielded disparate final growth rates despite identical strains and media (*E. coli* K-12 MG1655 in M9 glucose minimal media), 1.00±0.02 with 6 replicates and 0.79±.01 with 3 replicates, respectively. The only differences between the experiments were three experimental parameters: batch culture volumes (250 mL vs. 25 mL), optical densities when passed (variable vs. OD_{600nm} 1.2), and passage sizes (variable vs. 800µL) in the Charusanti et al. (18) and the LaCroix et al. (7) studies, respectively. ALEsim was constrained to allow only the jumps in growth rates observed in these studies and then simulated the expected fitness trajectories for the two different experimental parameters. The only differences explicitly defined in ALEsim were the different batch culture volumes, passage optical densities, and passage volumes. The results showed that

335 the difference in the final growth rates achieved can be sufficiently explained by the differences
336 in these parameters only (Figure 5). Furthermore, when simulating a legacy dataset for evolving
337 *E. coli* on glycerol minimal media, ALEsim was able to successfully predict that all experiments
338 (n=4) should reach fitness state 3 for the given experimental parameters, as reported in the study
339 (14). The largely different outcome in fitness (i.e., no fitness jumps vs. a significant increase) on
340 glucose, as well as a consistent prediction of fitness on a legacy glycerol dataset, further
341 highlights the importance of properly designing an experiment and validates ALEsim and its
342 parameterization.

343 **ALEsim Applications**

344 Simulations of ALE experiments with the derived beneficial mutation rate and fitness states can
345 enable statements to be made about optimality. The time required to see a given increase in
346 fitness was simulated for a range of increases in growth rate over a range of passage sizes
347 (Figure 6). The results show the average length of time needed to see a measurable change in
348 growth rate due to a beneficial mutation for a range of passage sizes. Figure 6 was derived for
349 growth rate increases that occur from a single mutational event. Based on the passage size and
350 length of time with no increase in growth rate, a conclusion about how close a population is to
351 reaching another state of increased fitness. For example, if a given evolution experiment has
352 achieved a certain growth rate, μ, and has not shown an increase in growth rate with a passage
353 size of 0.1% for 13 days, then there is no likely increase in growth rate available which is greater
354 than 0.10 hr^{-1} from a single mutational event.

355 Increasing the passage size raises the probability of capturing a beneficial mutation however this
356 also leads to an inflation in the resources needed to sustain the experiment (Figure 6). For
357 example, if an ALE experiment with a passage size of 0.1% were being passed twice a day

(every 12 hours), the same experiment with a passage size of 10% would need to be passed 6 times per day (every 4 hours). The magnitude of resources needed to maintain an experiment tend to scale with each batch. Thus, the more batches needing to be processed, the more media, pipette tips, culture vessels, and labor costs are required. A single person can feasibly do an experiment passed every 12 hours whereas passing every 4 hours would require coordinated effort by multiple persons or an automated platform. Therefore, understanding what is gained with the larger passage size is important before committing to such a large expenditure of resources. ALEsim can quantify the gains or losses achievable with different passage sizes to help identify the ideal experimental setup (Figure 6).

Mutation Frequency Analysis by Passage Size

Clones from the endpoint populations of each independent experiment were isolated and resequenced. Two clones showed hypermutating tendencies. This was identified by the number of mutations ($p<0.01$) and the presence of a mutation in *mutY* or *mutL*. Experiments with larger passage size led to an increase in the number of mutations found. Mutated alleles were therefore grouped by passage size. Clones isolated from larger passage size experiments, on average, had more alleles being selected (Figure 7). Of all mutations identified, those in *glpK* were specifically tracked. Mutations in *glpK* have previously been shown to be causal (with a significant impact on fitness) as well as ubiquitous, mutating more than any other alleles under glycerol growth conditions (14). Thus *glpK* is a good indicator of the how effective the various passage sizes are at fixing beneficial mutations. Consequently, there is a positive relationship between the fixing of *glpK* mutations and the passage size until saturation is reached. With the

passage size dropped to the lowest value (0.001%), the observed fraction that fixed was only 0.33 (2/6).

Applied and Environmental Microbiology

382 **Discussion**

383 The conceptual purpose of an ALE experiment is to move an organism towards a more optimal
384 (fit) state in the presence of a selection pressure. Absolute optimality is difficult, if even possible,
385 to define. It has been shown that even for a laboratory evolution, there is still room for evolution
386 after 50,000 generations (38). The continual ability of organisms to evolve and innovate makes it
387 difficult to analyze the results of an ALE experiment in the context of optimality. What is
388 immediately apparent is that there are diminishing returns. As an ALE experiment progresses,
389 the increase in growth rate or fitness tends to decrease in magnitude (1, 39-43). The smaller
390 increases take longer lengths of time to occur and become fixed in the population
391 (Supplementary Text). Given this property and the desire to understand and leverage the ALE
392 process, ALEsim was built and validated through performing a control experiment. ALEsim was
393 first parameterized with a set of control experiments using different passage sizes.
394 Parameterization revealed a beneficial mutation rate of $10^{-6.9}$-$10^{-8.4}$ mutations per cell division,
395 consistent with previously reported values and distinct fitness states (27, 28). Validation was
396 then carried out using additional legacy experiments and ALEsim proved sufficient for
397 explaining the differences in observed experimental outcomes (i.e., growth rates) based on the
398 parameters employed in each study (i.e., passage size, passage OD, and culture volume) (Figure
399 5). Lastly, ALEsim was applied to quantify tradeoffs in experimental design considerations for
400 desired outcomes and was used to demonstrate how it can be leveraged for determining the key
401 aspect of experiment termination.

402 The ability to optimize and design ALE experiments is possible with the ALEsim computational
403 framework. Given a certain amount of resources, ALEsim can calculate how best to deploy them
404 at different stages of an experiment to shorten project timelines and achieve desired outputs. For

example, near the beginning of the ALE experiments, the increases in growth rates found are typically quite large. Because of this, a large passage size does not have an additional benefit. This is evident in the experiment performed here in that passage sizes of 0.1%, 1%, and 10% mostly reached states 1, 2, and 3A at about the same time (Figure 3). In planning future ALE experiments, the added resource usage needed to maintain an experiment at a 10% passage size does not appear to be justified. However, the added benefits become apparent when looking at the transition from state 3A to 3B. It could then be suggested that if the goal is to get as close to the absolute optimal state as reasonably possible, the added resources of maintaining a 10% passage size experiment only need to be maintained after initial large increases in growth rate or fitness are found. This would not eliminate the difficulty in maintaining such an experiment, but would at least reduce the length of time the experiment would need to be run at such a high resource 'burn' rate. With ALEsim, these types of resource/fitness tradeoff analyses can now be calculated and should be leveraged in experimental design. The approach of dynamic resource allocation opens the door for project optimization typical of engineering process design.

Knowing the distance to optimality can aid in determining when to terminate an ALE experiment. The typical method of determining when to stop an ALE experiment is to subjectively determine that no more increases in fitness are being observed. However, this approach of waiting to observe a plateau in fitness can be artificial given a small passage size. An example of how this approach can be misleading is the observation that passage sizes of 0.1% and 1% showed no increase in growth rate after reaching state 3A for at least 15 days (Figure 3). However, given that slight increases in growth rates beyond state 3A to state 3B with a passage size of 10% were observed, it can be concluded that state 3A is not the optimal state. Thus, if only a 1% passage size was used, the experiment could be terminated before finding

state 3B. Further, it would be incorrect to compare experiments with a 10% passage size to a 1% passage size without understanding the context of the effects of the different passage sizes. Perhaps the best example of this is provided through the analysis of legacy ALE experiments (Figure 5). Two experiments with the same strain and media conditions yielded vastly different fitness outcomes. This difference is subsequently explainable within the scope of ALEsim. Therefore, having access to a computational framework such as ALEsim can enable the researcher to make an informed decision about when to terminate an experiment given the capacity and resources of the experimental setup and the desired/acceptable outcome. This type of termination analysis is laid out in Figure 6 and can be calculated *de novo* for any experiment given the current growth rate and passage size. It also should be noted that this type of analysis could result in a standard for the ALE community as one could state the ALEsim generated $\Delta\mu$ at the time of termination.

The ability to design and carry out complicated and high resource burn ALE experiments is likely only feasible though automation of the ALE process. Automation was utilized here and in previous studies (4, 7, 44). Manual processes are often hindered by researcher availability whereas machines can measure and pass around the clock (e.g., approximately 5-7 passages per day were performed in automated studies (4, 7, 44), compared to 1-2 per day manually (14, 15, 18). Thus, the ability to automate and optimize ALE is likely to accelerate adoption of the ALE experimental technique and broaden the application areas. Furthermore, the ALEsim framework and output can also be used as a basis for modeling much of the legacy data currently available for ALE experiments which include lag, exponential, stationary, and/or stressed phases. As the selection pressure in such experiments is more complex and growth is defined by more than the growth rate parameter (e.g. lag phase duration, stationary phase mutation rate, growth phase

transistions, etc…), ALEsim in its current format would have to be expanded. Nonetheless, ALEsim and it parameterization here demonstrates the utility of using simulated design in the ALE process and establishes a portable code base.

The field of adaptive laboratory evolution is expanding, largely due to lower costs of next generation sequencing. Innovative applications are appearing and are being applied to a range of organisms (1, 3). This growth in ALE use has occurred without a standard operating procedure for performing and quantifying these experiments. Consequently, this leads to ill-defined endpoints of experiments and the inefficient use of resources. The ALEsim computational platform developed here would provide a basis with which to quantify experiments and aid in their design; matching the desired outcome with resources available.

List of Supplementary Files

- Supplemental text (calculations)
    - Supplementary Table – ALE Fitness Stats
    - Supplementary Figure 1 – Sensitivity analysis of dry weight per cell values
- Supplementary File 1 – MATLAB m files: ALE Model m-files

Acknowledgements and Funding Information

This work was supported by the Novo Nordisk Foundation Center for Biosustainability

We'd like the Marc Abrams, Troy Sandberg, and Richard Szubin for their assistance with this manuscript.

Tables

None

Figure Legends

Figure 1 - ALEsim Flow Chart

A workflow outlining the logical steps the simulator takes when performing a single simulated ALE experiment. Due to the stochastic nature of ALE experiments, *in vivo* and *in silico*, multiple experiments are averaged together to identify general trends.

Figure 2 - Governing Equations, Assumptions, and Parameters for ALEsim

a) Microbe growth occurs according to an exponential growth curve where μ is the growth rate, t is the time elapsed, N_0 is the initial cell count at t=0, and N(t) is the cell count at a given time, t. No lag phase or stationary phase is modeled. The total cell count (N(t)) is determined by the summation of exponential growth curves for all individual cells lines. b) Favorable mutations occur during cell growth according to a binomial distribution where each cell division represents one Bernoulli trial with a probability of success equal to the beneficial mutation rate (BMR). c) Each flask is modeled as a completely homogenous culture. d) The number of cells represented for each cell line in each inoculum is randomly chosen according to a normal distribution with a mean and variance equal to the number of cells represented in the flask, N_{Green}^{Flask} times the ratio of the flask volume, V_{Flask}, to inoculum volume, $V_{Inoculum}$. e-g) The volume of media per flask, inoculum volume, and passage optical density can be altered. h) The simulated ALE experiment can be stopped after a specified amount of time or maximum number of flasks. i) Based on the

relative growth rate increases seen in ALE experiments, a range of allowable growth rate increases is determined. j) Based on matching the evolution trajectory (plot of growth rate vs. flask #) with varying the beneficial mutation rate (BMR), the probability of a favorable mutation is obtained. k) Since each ALE is based on randomly generated mutations, multiple ALE simulations are averaged together to get repeatable results from the same parameters. The number of simulations is user controlled.

Figure 3 – Fitness Trajectory of *E. coli* evolved on Glycerol

The absolute growth rates of independently evolved cultures of *E. coli* as fitted by a cubic spline for all ALE experiments separated by the different passage sizes. Dashed lines represent regions where the spline fit is based on sparse data, and therefore not considered accurate. The small upturn in growth rates at the endpoint is an artifact of the spline interpolation and is ignored when determining endpoint growth rates. All except five ALE experiments reached fitness State 3. The rate at which the final growth rate was achieved varied. The hypermutating strain with a passage size of 10% reached State 3 significantly faster than all others (it possessed a mutation in *mutY*). The purple hypermutating strain was identified as a potential hypermutating strain based on the number of mutations fixed (p=0.003, FDR=0.087) and the presence of a frame shift insertion in *mutL*.

Figure 4 – Distribution of Fitness Increases in Glycerol ALE

A histogram of the normalized increases in growth rate (μ_{max} = 0.64 hr^{-1}) attributed to each jump for the different experiments. The fitness increases were categorized by which state transition was made. The different passage sizes (indicated by different colors) did not show any significant variance in the ability to fix distinct increases in growth rate. A few small jumps not

517 shown are small observed increases in fitness that did not jump between any of the states
518 identified.

519 **Figure 5 – Simulated vs Experimental Results with Large and Small Passage Sizes**

520 Two ALE experiments of *E.coli* MG1655 in glucose M9 minimal media were simulated using
521 ALEsim. The strain and media conditions were identical in the two experiments. The only
522 differences were in the culture volume (25ml vs. 250mL), optical density when passed (variable
523 vs. 1.2 OD_{600nm}), and passage volume (variable vs 800µL). The variable nature of the optical
524 density when passed and the passage size in the latter experiment was a consequence of
525 manually passing the culture each day. The former experiment employed an automated system of
526 monitoring and passing the culture to maintain consistency. Despite being the same strain and
527 conditions, the final fitness achieved in the two experiments were quite different. ALEsim was
528 used to simulate these same experiments with the only differences being the three
529 aforementioned parameters. Consequently, the ALEsim results showed that the differences in
530 these parameters were sufficient to explain why the final growth rates achieved were different,
531 further highlighting the importance of choosing these parameters properly. The simulated
532 resulted are represented by a 95% confidence interval. The confidence interval for Experiment
533 #2 is too small to be visible.

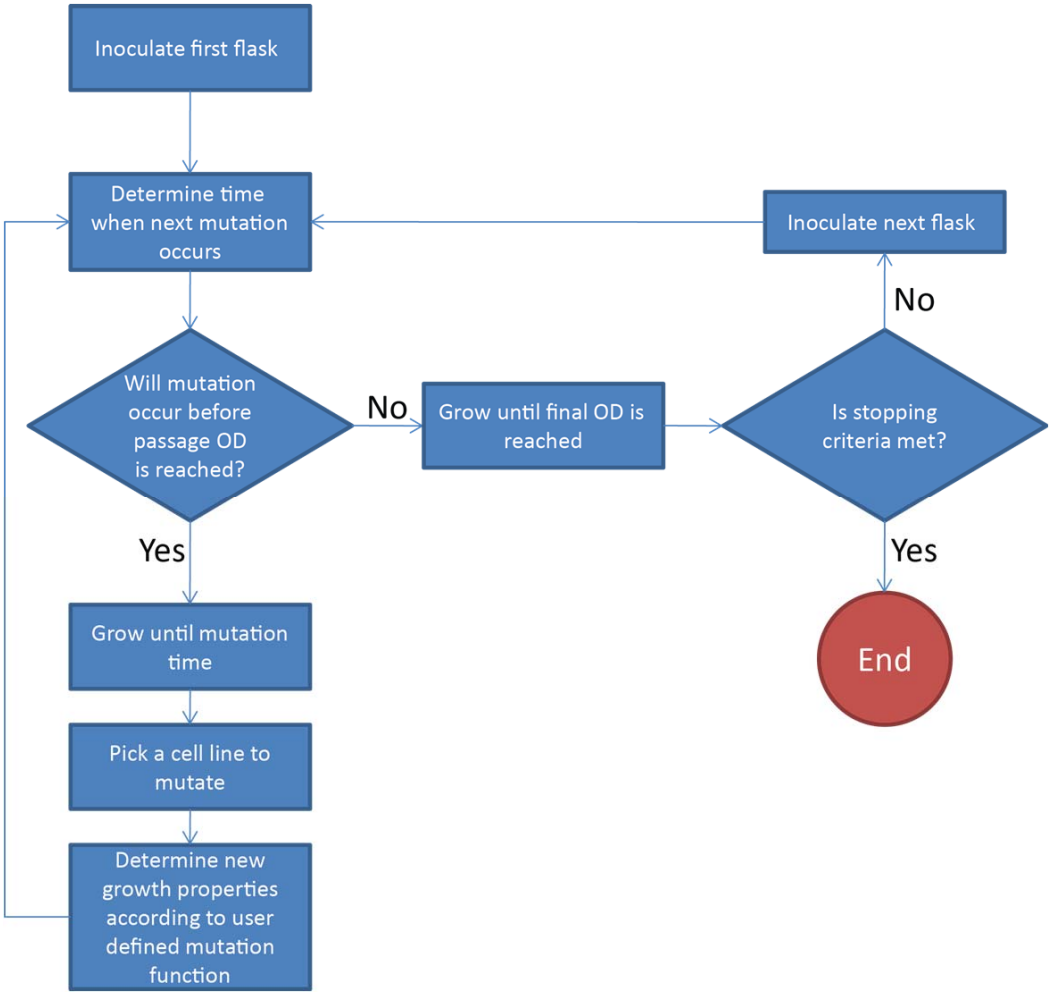
534 **Figure 6 – Upper Bound on possible jumps in growth rates**

535 A. Upper bounds on possible jumps in growth rates are shown. At a given point in time, a jump
536 that reaches above the upper bound is statistically infeasible (95% confidence) from a single
537 mutation, whereas jumps that stay below the line are possible. B. The upper bound on jumps is
538 shown for varying passage sizes. These experiments were simulated with parameters that

matched the experimental parameter used. Increasing the passage size can have a significant impact on the upper bound. Consequently, the time required to eliminate jumps of certain magnitudes can take much longer to achieve. However, as the passage size increases there comes a point when the returns begin to diminish such that passage sizes between 0.1% and 10% did not show a large difference in the time required to find a given jump. C. Relative amount of resources needed to perform an ALE experiment normalized to the lowest passage size. As the passage size is increased the resource usage begins to increase greatly.

Figure 7 – Genetic Analysis – By Passage Size

A bar chart representing the observed fraction of mutations at a given passage volume. As a general trend, the larger the passage size, the greater the probability of a mutation in a given allele fixing in the population. A key mutation in the *glpK* gene is displayed as well as all mutations. The ordinal rank of passage size was compared to the observed fraction of mutations using a Wilcoxon rank test and resulted in p-values of 0.008 and 0.024 for all mutations and *glpK* mutations, respectively.

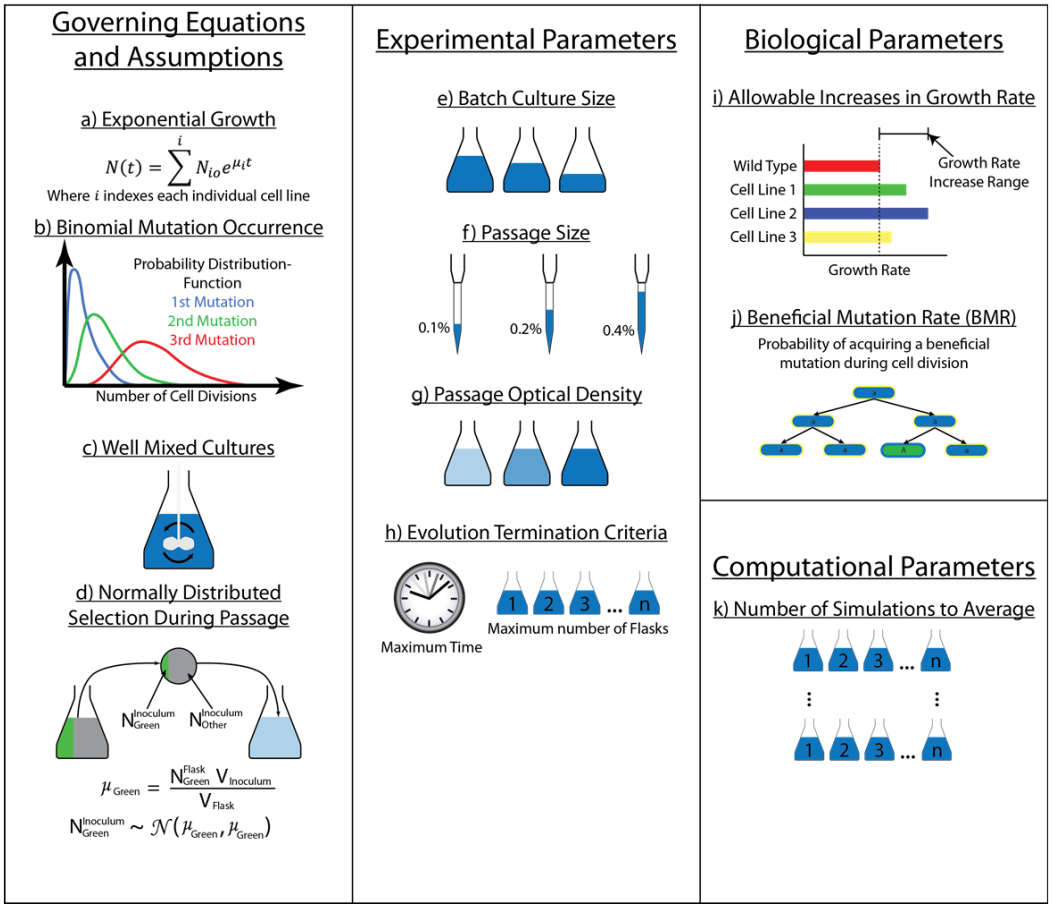


554

555 **Figure 1**

556

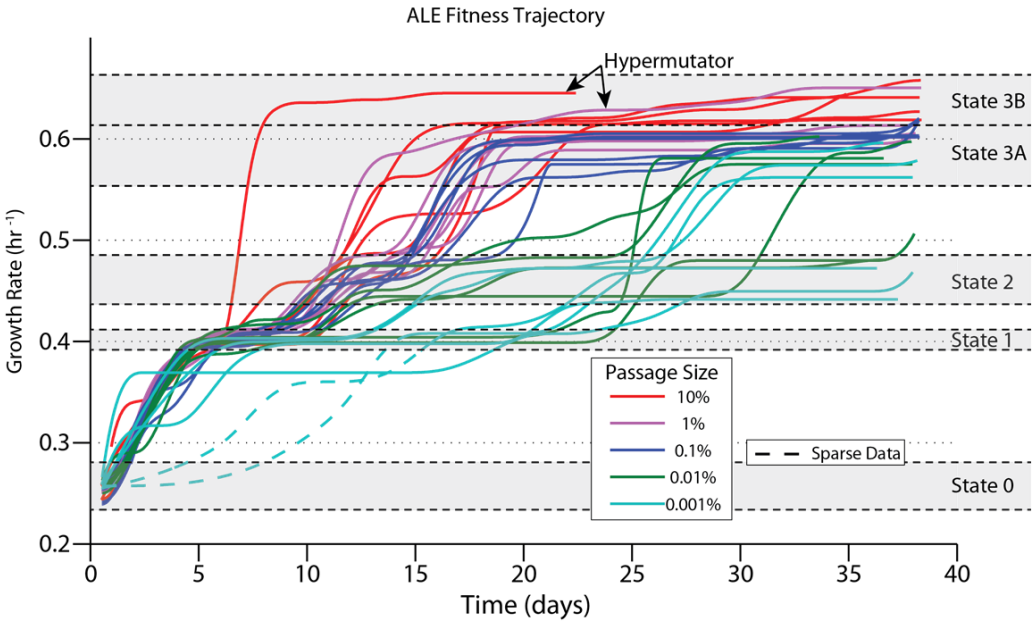
557



558

559 **Figure 2**

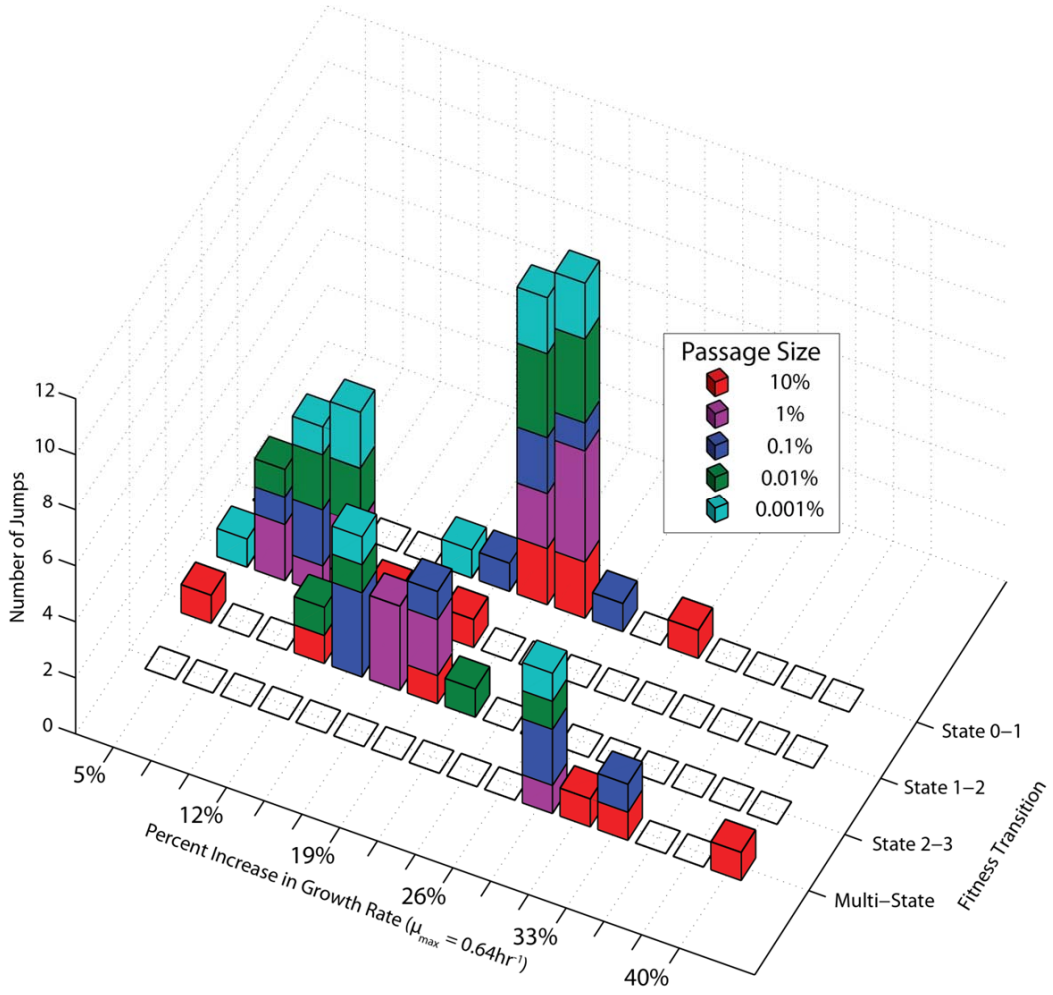
560



561

562 **Figure 3**

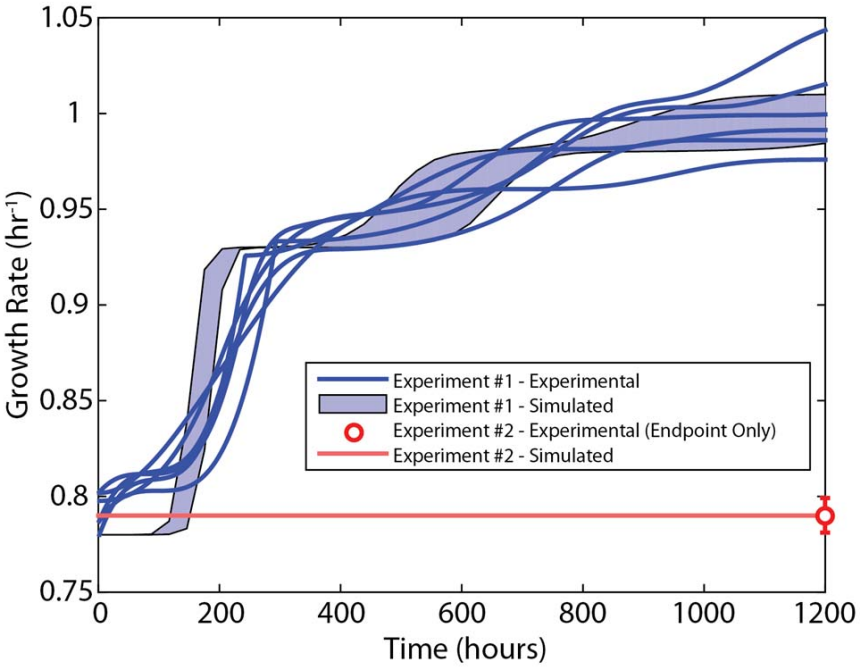
563



564

565 **Figure 4**

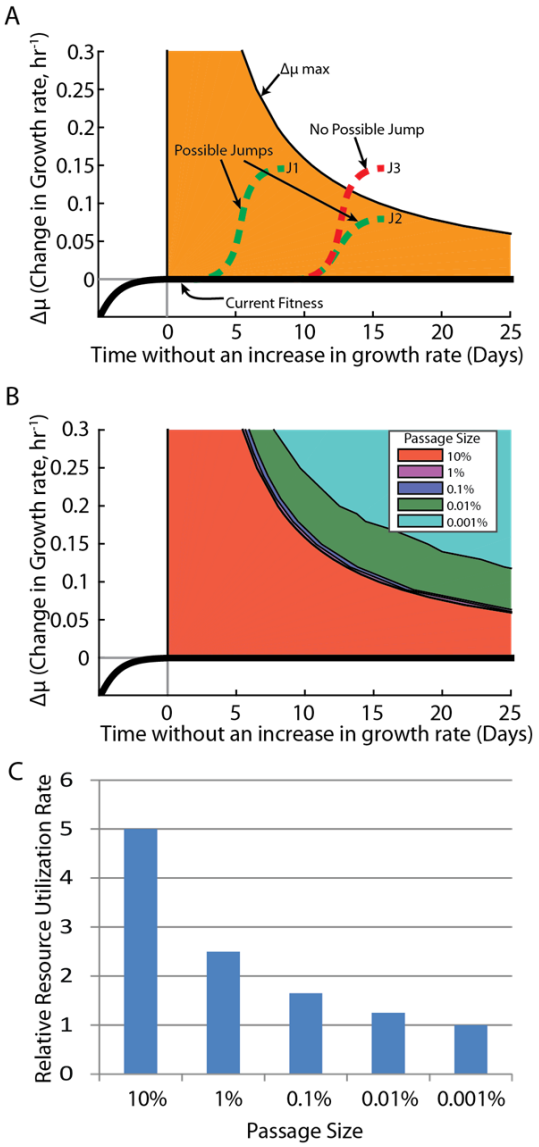
566



567

568 **Figure 5**

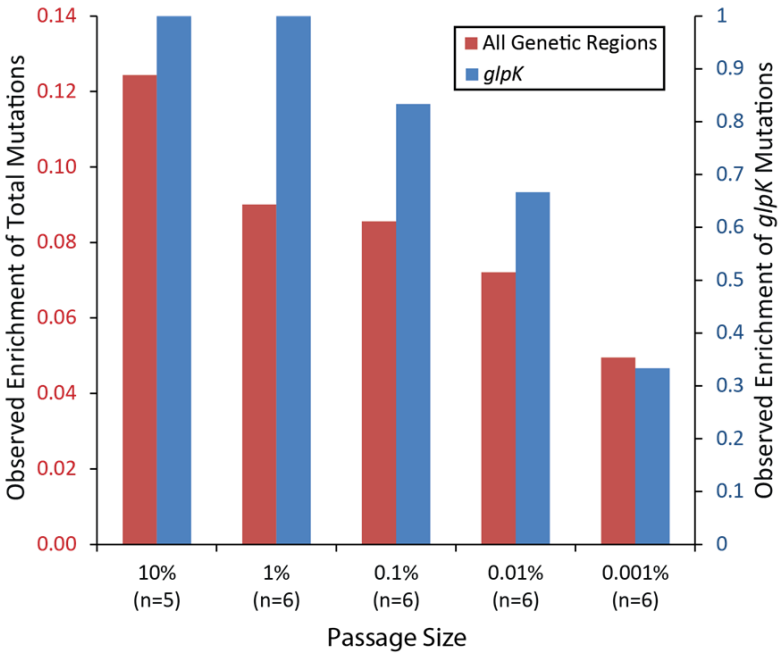
569



570

571 **Figure 6**

572



573

574 **Figure 7**

References

1. **Palsson B.** 2010. Adaptive laboratory evolution. Microbe.
2. **Conrad TM, Lewis NE, Palsson BO.** 2011. Microbial laboratory evolution in the era of genome-scale science. Mol Syst Biol **7:**509.
3. **Dragosits M, Mattanovich D.** 2013. Adaptive laboratory evolution -- principles and applications for biotechnology. Microb Cell Fact **12:**64.
4. **Sandberg TE, Pedersen M, LaCroix RA, Ebrahim A, Bonde M, Herrgard MJ, Palsson BO, Sommer M, Feist AM.** 2014. Evolution of Escherichia coli to 42 degrees C and subsequent genetic engineering reveals adaptive mechanisms and novel mutations. Mol Biol Evol **31:**2647-2662.
5. **Harcombe WR, Delaney NF, Leiby N, Klitgord N, Marx CJ.** 2013. The ability of flux balance analysis to predict evolution of central metabolism scales with the initial distance to the optimum. PLoS Comput Biol **9:**e1003091.
6. **Gresham D, Hong J.** 2015. The functional basis of adaptive evolution in chemostats. FEMS Microbiol Rev **39:**2-16.
7. **LaCroix RA, Sandberg TE, O'Brien EJ, Utrilla J, Ebrahim A, Guzman GI, Szubin R, Palsson BO, Feist AM.** 2015. Use of adaptive laboratory evolution to discover key mutations enabling rapid growth of Escherichia coli K-12 MG1655 on glucose minimal medium. Appl Environ Microbiol **81:**17-30.
8. **Bacun-Druzina V, Cagalj Z, Gjuracic K.** 2007. The growth advantage in stationary-phase (GASP) phenomenon in mixed cultures of enterobacteria. FEMS Microbiol Lett **266:**119-127.
9. **Vasi F, Travisano M, Lenski RE.** 1994. Long-term experimental evolution in Escherichia coli. II. Changes in life-history traits during adaptation to a seasonal environment. American Naturalist**:**432-456.
10. **Bachmann H, Fischlechner M, Rabbers I, Barfa N, Branco dos Santos F, Molenaar D, Teusink B.** 2013. Availability of public goods shapes the evolution of competing metabolic strategies. Proc Natl Acad Sci U S A **110:**14302-14307.
11. **Raynes Y, Halstead AL, Sniegowski PD.** 2014. The effect of population bottlenecks on mutation rate evolution in asexual populations. J Evol Biol **27:**161-169.
12. **Campos PR, Wahl LM.** 2010. The adaptation rate of asexuals: deleterious mutations, clonal interference and population bottlenecks. Evolution **64:**1973-1983.
13. **Campos PR, Wahl LM.** 2009. The effects of population bottlenecks on clonal interference, and the adaptation effective population size. Evolution **63:**950-958.
14. **Herring CD, Raghunathan A, Honisch C, Patel T, Applebee MK, Joyce AR, Albert TJ, Blattner FR, van den Boom D, Cantor CR, Palsson BO.** 2006. Comparative genome sequencing of Escherichia coli allows observation of bacterial evolution on a laboratory timescale. Nat Genet **38:**1406-1412.
15. **Ibarra RU, Edwards JS, Palsson BO.** 2002. Escherichia coli K-12 undergoes adaptive evolution to achieve in silico predicted optimal growth. Nature **420:**186-189.
16. **Lee DH, Palsson BO.** 2010. Adaptive evolution of Escherichia coli K-12 MG1655 during growth on a Nonnative carbon source, L-1,2-propanediol. Appl Environ Microbiol **76:**4158-4168.
17. **Conrad TM, Joyce AR, Applebee MK, Barrett CL, Xie B, Gao Y, Palsson BO.** 2009. Whole-genome resequencing of Escherichia coli K-12 MG1655 undergoing short-term laboratory

evolution in lactate minimal media reveals flexible selection of adaptive mutations. Genome Biol **10:**R118.

18. **Charusanti P, Conrad TM, Knight EM, Venkataraman K, Fong NL, Xie B, Gao Y, Palsson BO.** 2010. Genetic basis of growth adaptation of Escherichia coli after deletion of pgi, a major metabolic gene. PLoS Genet **6:**e1001186.
19. **Wahl LM, Gerrish PJ.** 2001. The probability that beneficial mutations are lost in populations with periodic bottlenecks. Evolution **55:**2606-2610.
20. **Hubbarde JE, Wahl LM.** 2008. Estimating the optimal bottleneck ratio for experimental evolution: the burst-death model. Math Biosci **213:**113-118.
21. **Deatherage DE, Barrick JE.** 2014. Identification of mutations in laboratory-evolved microbes from next-generation sequencing data using breseq. Methods Mol Biol **1151:**165-188.
22. **Neidhardt FC, Ingraham JL, Schaechter M.** 1990. Physiology of the bacterial cell : a molecular approach. Sinauer Associates, Sunderland, Mass.
23. **Pramanik J, Keasling JD.** 1997. Stoichiometric model of Escherichia coli metabolism: incorporation of growth-rate dependent biomass composition and mechanistic energy requirements. Biotechnol Bioeng **56:**398-421.
24. **Kibota TT, Lynch M.** 1996. Estimate of the genomic mutation rate deleterious to overall fitness in E. coli. Nature **381:**694-696.
25. **Fogle CA, Nagle JL, Desai MM.** 2008. Clonal interference, multiple mutations and adaptation in large asexual populations. Genetics **180:**2163-2173.
26. **Tenaillon O, Rodriguez-Verdugo A, Gaut RL, McDonald P, Bennett AF, Long AD, Gaut BS.** 2012. The molecular diversity of adaptive convergence. Science **335:**457-461.
27. **Desai MM, Fisher DS, Murray AW.** 2007. The speed of evolution and maintenance of variation in asexual populations. Curr Biol **17:**385-394.
28. **Perfeito L, Fernandes L, Mota C, Gordo I.** 2007. Adaptive mutations in bacteria: high rate and small effects. Science **317:**813-815.
29. **Gerrish PJ, Lenski RE.** 1998. The fate of competing beneficial mutations in an asexual population. Genetica **102-103:**127-144.
30. **Uecker H, Hermisson J.** 2011. On the fixation process of a beneficial mutation in a variable environment. Genetics **188:**915-930.
31. **Lande R.** 2007. Expected relative fitness and the adaptive topography of fluctuating selection. Evolution **61:**1835-1846.
32. **Wright S.** 1929. Fisher's Theory of Dominance. American Naturalist **63:**274-279.
33. **Haldane JBS.** 1927, p 838-844. Mathematical Proceedings of the Cambridge Philosophical Society.
34. **Fisher RA.** 1930. The genetical theory of natural selection. The Clarendon press, Oxford,.
35. **Wahl LM, Zhu AD.** 2015. Survival probability of beneficial mutations in bacterial batch culture. Genetics **200:**309-320.
36. **Foster JA.** 2001. Evolutionary computation. Nat Rev Genet **2:**428-436.
37. **Reyes LH, Almario MP, Winkler J, Orozco MM, Kao KC.** 2012. Visualizing evolution in real time to determine the molecular mechanisms of n-butanol tolerance in Escherichia coli. Metab Eng **14:**579-590.
38. **Wiser MJ, Ribeck N, Lenski RE.** 2013. Long-term dynamics of adaptation in asexual populations. Science **342:**1364-1367.
39. **Kryazhimskiy S, Rice DP, Jerison ER, Desai MM.** 2014. Microbial evolution. Global epistasis makes adaptation predictable despite sequence-level stochasticity. Science **344:**1519-1522.
40. **Khan AI, Dinh DM, Schneider D, Lenski RE, Cooper TF.** 2011. Negative epistasis between beneficial mutations in an evolving bacterial population. Science **332:**1193-1196.

41. **Chou HH, Chiu HC, Delaney NF, Segre D, Marx CJ.** 2011. Diminishing returns epistasis among beneficial mutations decelerates adaptation. Science **332:**1190-1192.
42. **Barrick JE, Kauth MR, Strelioff CC, Lenski RE.** 2010. Escherichia coli rpoB mutants have increased evolvability in proportion to their fitness defects. Mol Biol Evol **27:**1338-1347.
43. **Perfeito L, Sousa A, Bataillon T, Gordo I.** 2014. Rates of fitness decline and rebound suggest pervasive epistasis. Evolution **68:**150-162.
44. **Sandberg TE, Long CP, Gonzalez JE, Feist AM, Antoniewicz MR, Palsson BO.** 2016. Evolution of E. coli on [U-13C]Glucose Reveals a Negligible Isotopic Influence on Metabolism and Physiology. PLoS One **11:**e0151130.